International Journal of Finance and Managerial Accounting, Vol.7, No.26, Summer 2022 IJFMA

A Comparative Study of the Performance of Stock Trading Strategies Based on LGBM and CatBoost Algorithms

Seyed Majid Mousavi Anzahaei
Ph.D. Candidate, Department of Financial Management, Faculty of Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran
Mousavi.Smajid@gmail.com

Hashem Nikoomaram
Professor of financial management, Department of Financial Management, Faculty of Management and Economics, Science and Research Branch, Islamic Azad University, Tehran, Iran
(Corresponding Author)
nikoomaram@srbiau.ac.ir

Submit: 20/01/2021 Accept: 27/02/2021

ABSTRACT
Today, investment in the stock market requires novel and efficient methods along with effective trading strategies for more accurate prediction of stock price future movements. This paper compares the performance of implementing LGBM and CatBoost trading strategies on a portfolio, which is formed, based on fundamental analysis and future study. First with use of future study and expert's opinion, stock market scenarios designed and a portfolio consist of 6 fundamental stocks is built. In next step for each selected stocks a model is developed by means of LGMB and CatBoost algorithms and related stocks data from 2014 to 2019 to predict stock price movement. Model inputs includes, technical indicators, stocks trading data and some market and fundamental index. Bayesian hyper parameter was used to optimize the model's key parameters. Results show that models optimized with Bayesian hyper parameter are more accurate than models, which optimized with grid search and implementing short-term trading strategies based on gradient boosting machine (LGBM) prediction signals cause better performance in comparison with CatBoost based strategies and Tehran Stock Exchange Index.

Keywords:
Trading Strategies, Future Study, Light Gradient Boosting Machine (LGBM), Categorical boosting (CatBoost), Bayesian hyper parameter.

1. Introduction

According to efficient market Hypothesis (EMH), all information is available and continuously processed by the market and all historical, general and private information about an asset is embedded in to its current price and it is not possible to outperform the market (Fama&Bolom,1966).However, investment managers in active portfolio management have always sought to find a way to outperforming market index return and the buy and hold strategy. According to recent research, the financial market does not behave completely random and it is possible to predict its changes, Atsalakis (2009).

The three key elements of any successful investment are the stock price forecast, which states what decision the investor should make (buy or sell), the timing of the transaction, and the amount of investment. (Noorbakhsh and Tehrani, 1388)

Therefore, predicting the trend of stock price movement makes it possible to use appropriate strategies for trading and trading strategies also examine the optimal entry points with short-term, medium-term and long-term perspectives.

Active portfolio management requires the use of appropriate trading strategies and the adoption of an appropriate strategy requires analysis on stocks.

There are two main philosophies in stock trading, fundamental analysis and technical analysis (Shynkevich, et al., 2017). Traditional technical analysts have developed many indices and sequential analytical methods that may reflect the trends in the movements of the stock price. Meanwhile, the prices are affected by many macro-economic factors, fundamental factors of companies and the involvement of public investors. Therefore, some criticism of technical analysis is that it only considers transactional data of stocks and completely ignores the fundamental factors of companies which might be helpful, if the market is in weak- form efficiency. (Zhang, et.al, 2018)

Although technical models have been widely used in stock market prediction, but studies show that these methods have not been successful enough.

With the advancement of science, researchers have used time series models and artificial neural networks for better predictions, Lawrence (1997).

Today, technical indicators are widely used as inputs to machine learning prediction algorithms, Atsalakis (2009). Financial forecasting based on computational intelligence approaches often uses technical analysis (TA) to form features used as inputs to the approaches. Time series of stock price and trading volume are utilized to compute a technical indicator (TI) where a composition of open, low, high and close price values and volume size is taken over a certain time period, (shynkevich, et.al, 2017).

Many studies show that individual algorithms cannot solve problems effectively according to funding of researchers such as (Nanni& Lumini, 2009), Ensemble methods have better performance in prediction. Ensemble methods can be roughly categorized into two groups according to their structure: Parallel and sequential. Some research indicates the superiority of Parallel methods, (Brwon & Mues, 2012) and others to sequential methods, (Yufi et.al, 2017). Ensemble learning combines multiple algorithms that process different hypotheses to form a better hypothesis, thus making good predictions, (Nascimento, Coelho, & Canuto, 2014).

According to the study of Nanni and Lumini (2009), and Lessmann et al. (2015) , ensemble methods perform better than single AI and statistical methods. The promising results encourage further exploration of its structures, combination strategies, and mechanisms.

Ensemble methods can be roughly categorized into two groups according to their structures: parallel and sequential (Duin &Tax, 2000). The parallel ensemble is the combination of different learning algorithms, each of which generates an independent model in parallel. By contrast, the first algorithm in the sequential ensemble learns to generate a model, and then the second algorithm learns to correct the former model.

Researchers have mainly focused on parallel ensemble methods, (Paleologo, Elisseeff, & Antonini, 2010; Wang, Ma, Huang, & Xu, 2012). Unlike the bagging algorithm that fits the base models in parallel, the boosting approach sequentially builds models. The basic idea of boosting is combining a series of weak base learners, which are normally regression trees, into a strong one. The weak learner herein refers to a model that only performs slightly better than a random guess. Boosting fits additive base learners to minimize the loss function provided. Loss function measures how well the model fits the current data. The process of boosting continues until the loss function reduction becomes limited, (Yufei Xia, et.al.2017).

For a long time, the tuning of hyperparameters for learning algorithms was solved by simple exhaustive methods such as grid search guided by cross-validation error. Grid search does work in practice, but it suffers from serious drawbacks such as a search space complexity that grows exponentially with the number of hyperparameters tuned.(Julian, levesque, et.al, 2020) Therefore, how to make the automatic tuning algorithm achieve high precision and high efficiency has always been a problem that has not yet been fully solved in machine learning. In this regard, Bayesian optimization is a powerful tool for finding the optimum point of objective functions that are unknown and expensive to evaluate. (Parsa, et.al, 2020)

Many researches have been done in the field of active stock portfolio management with the help of machine learning algorithms, such as Gholamian and Davoodi (1396), Fallahpour and Dana (1395), Moshari et al. (1398), Choudry and Gary (2008), Nani and Lumi (2009), Yufi et al. (2017), Patel et al. (2015), Zhang et al. (2018), Gulin et al. (2017), Sakar (2019).

In Table 1, we review some of the research conducted on the application of trading strategies resulting from stock price prediction

Table 1. Summary of researches

Evaluation Criteria	Prediction Variable	Sample	Learning Algorithm	Input Variable	Year	Researchers
Strategy Return	Stock Price	18 Filtered Stock	FUZZY - GA	16 Technical	2015	Raei& Hoseini
Strategy Return	Stock Price	6 Random Stock	KNN-GA ANN-GA	22 Technical	2015	Tehrani,et.al
Strategy Return	Stock Price Trend	10 Random Stock	Weighted SVM	5 Technical	2016	Bajelan, et.al
Strategy Return	Stock Price Trend	The Best 15 Stock	MLP+ (ACOR GA, PSO)	Technical Trading Rule	2020	Saranj, et.al
Model Return	Price Trend	50 Random Stock	SVM, ANN, KNN	10 Technical	2017	Shynkevch, et.al,
Model performance	Stock Price Trend	42 Digital Currency	RF, SVM, LGBM	40 Technical	2020	Xiaolei , et.al.
Strategy Return	Stock Price Trend	Scenario Based Portfolio	CatBoost and LGBM (Optimized with Bayesian Hyper parameter)	14 technical 1 Tabloo 2 Market 2 Fundamental 2 Statistical	2020	This Research

There are a variety of prediction methods in Economic and Finance Literature. These techniques were classified into four groups based on the type of tool and the type of data used:

1) Technical Analysis
2) Fundamental Analysis Methods
3) Traditional Time Series Methods (Econometrics)
4) 4-Machine Learning

As mentioned, one of the key factors in applying trading strategies is to conduct stock price analysis and predict its future trend.

In general, stock trading strategies can be summarized as Figure 1.

Hyperparameter Tuning (optimization):

Hyperparameters are different from the internal model parameters, such as the neural network's weights, which can be learned from the data during the model-training phase. Before the training phase, we would like to find a set of hyperparameter values, which archive the best performance on the data in a reasonable amount of time. This process is called hyperparameter optimization or tuning. It plays a vital role in the prediction accuracy of machine learning algorithms, (Jia.Wue,et.al, 2019)

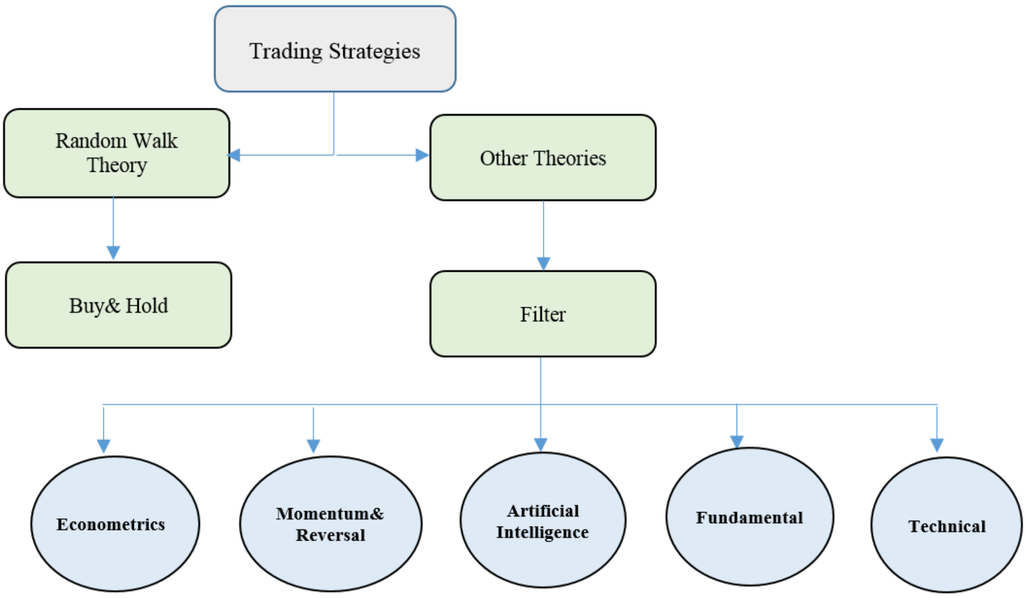

Figure1: Different types of Trading Strategies in stock market (Researcher Findings, 1399)

2. Grid Search

To overcome the drawbacks of manual search, automatic search algorithms have been proposed, such as grid search or Cartesian hyperparameter search the principle of grid search is exhaustive searching. Grid search trains a machine-learning model with each combination of possible values of hyperparameters on the training set and evaluates the performance according to a predefined metric on a cross validation set. Finally, grid search outputs hyperparameters that achieve the best performance. Although this method achieves automatic tuning and can theoretically obtain the global optimal value of the optimization objective function, it suffers from the curse of dimensionality, i.e., the efficiency of the algorithm decreases rapidly as the number of hyperparameters being tuned and the range of values of hyperparameters increase. (Jia.Wue,et.al, 2019)

3. Bayesian Hyper Parameter:

Bayesian optimization is an effective method for solving functions that are computationally expensive to find the extrema (Brochu, et.al,2010) It can be applied for solving a function, which does not have a closed-form expression. It can also be used for functions, which are expensive to calculate, the derivatives are hard to evaluate, or the function is non-convex. In this paper, the optimization goal is to find the maximum value at the sampling point for an unknown function *f*:

$$x+ = \arg\max_{x \in A} f(x) \qquad (1)$$

Where *A* denotes the search space of *x*. Bayesian optimization derives from Bayes' theorem i.e., given Evidence data *E*, the posterior probability *P* (*M*|*E*) of a model *M* is proportional to the likelihood *P* (*E*|*M*) of Overserving *E* given model *M* multiplied by the prior probability of *P* (*M*):

$$P\ (M\ jE) / P\ (EjM)\ P(M) \qquad (2)$$

(Jia.Wue,et.al, 2019)

Light Gradient Boosting (LGBM)

The LightGBM algorithm is introduced in detail. LightGBM is a novel GBDT (Gradient Boosting Decision Tree) algorithm, proposed by Ke and colleagues in 2017, which has been used in many different kinds of data mining tasks, such as classification, regression and ordering (Ke et al., 2017). The LightGBM algorithm contains two novel techniques, which are the gradient-based one-side sampling and the exclusive feature bundling, respectively.

Given the supervised training set $X = \{(xi, yi)\}^{n}{}_{i}= 1$, *minimizes the expected value of a specific loss function*

L (y, f(x)) as follows: *(3)*

$f^{\wedge}$ = arg min$E\ L\ (y, f\ (x))$, (Xiaolei, et.al, 2018)

Categorical boosting (CatBoost)

CatBoost is a new gradient boosting decision tree (GBDT) algorithm that can handle categorical features well. This algorithm is different from traditional GBDT algorithms in the following aspects :(1) Dealing with categorical features during training time instead of preprocessing time. CatBoost allows the use of whole dataset for training. According to Prokhorenkova et al. (2018), target statistics (TS) is a very efficient method for handing categorical features with minimum information loss. Specifically, for each example, CatBoost performs a random permutation of the dataset and computes an average label value for the example with the same category value

Placed before the given one in the permutation. If a permutation is (4)

$\Theta = [\sigma_1, \cdots, \sigma_n]_n^T$, it is substituted with:

$$x_{\sigma_{p,k}} = \frac{\sum_{j=1}^{p-1} [x_{\sigma_{j,k}} = x_{\sigma_{p,k}}] \cdot Y_{\sigma j} + \beta \cdot P}{\sum_{j=1}^{p-1} [x_{\sigma_{j,k}} = x_{\sigma_{p,k}}] + \beta}$$

Where P is a prior value and is the weight of the prior. For regression tasks, the standard technique for calculating prior is to take the average label value in the dataset.(Guomin , Huang,et.al, 2019)

4. Methodology

In this research, we try to compare the performance of trading strategies based on the predictions of LGBM and CatBoost learning algorithms with each other and buying and Hold strategies.

The present study is applied research in terms of purpose, and in terms of methodology, it is quasi-experimental with use of modeling.

Data were collected using computer databanks and Rahavard Novin 3 Software and by referring to the library of the stock exchange and Codal Website belonging to the stock exchange market

In addition, financial statements of firms, including balance sheets, cash flow statements, and notes attached to financial statements at the end of each financial year (March 19th), were applied as data collection tools.

Research population included all companies listed on the Tehran Stock Exchange during a 5-year period (2015-2019). Sample was selected using systematic elimination method, encompassing companies that met the following criteria:

1) with a free float above 20%
2) At least 60% of trading days, their shares are traded.
3) Selected from future study for 2 years horizon.
4) 4-With liquidity ratio above 90%
5) 5-Have fundamentally better conditions in the relevant industry

Data related to selected sample was collected for the period 2014/03/21 to 2019/03/18

In this study, PYTHON software for modeling, SPSS 26 for Statistical analysis, MicMac and Scenario Wizard for future study are used. Also, technical library in www.ta_lib.org is used for calculation of Technical indicators. Input variable include 14 technical indicators, 6 stock trading data (Trading volume, first price, closed price, highest, lowest and the last price) ,5 statistical and 5 market and fundamental data according to Table 2. Finally, for each data point, there is a matrix consisting of 30 input variables and an output variable in two statues, "1" (Ascending) and "-1" (Descending) Which shows the trend of price movement in the next 15 days. In general, the present research has been carried out during four Executive phases according to Figure 1:

Table2. Input Valuables (Technical indicators & Stock Trading Data)

Symbol	Index	Symbol	Index
%R	Williams %R	Stock Trading	
CCI	Commodity Chanel Index	PO	Opening
MOM	Momentum Indicator	PL	Lowest
TSF	Time Series Forecast	PH	Highest
OBV	On- Balance Volume	PL	Last
MFI	Money Flow Index	PC	Closing

Symbol	Index	Symbol	Index
Statistical		Vol.	Volume
STDDEV	Standard Deviation	Technical Indicator	
Volume 7,14,20,30	Trading Volume 7,14,20&30 days	SMA	Moving Average
Market		EMA	Exponential Moving Average
TINDX	Tehran Overall Index	ADX	Average Directional Moving Index
WINDX	Tehran Weight Index	D%	Stochastic (D %)
Fundamental & Tabloo		K%	Stochastic (K %)
P/E	Price/ EPS	MACD	MACD
P/S	Price/ Sales	SAR	Parabolic SAR
PB	Power of Buyer	RSI	Relative Strength Index

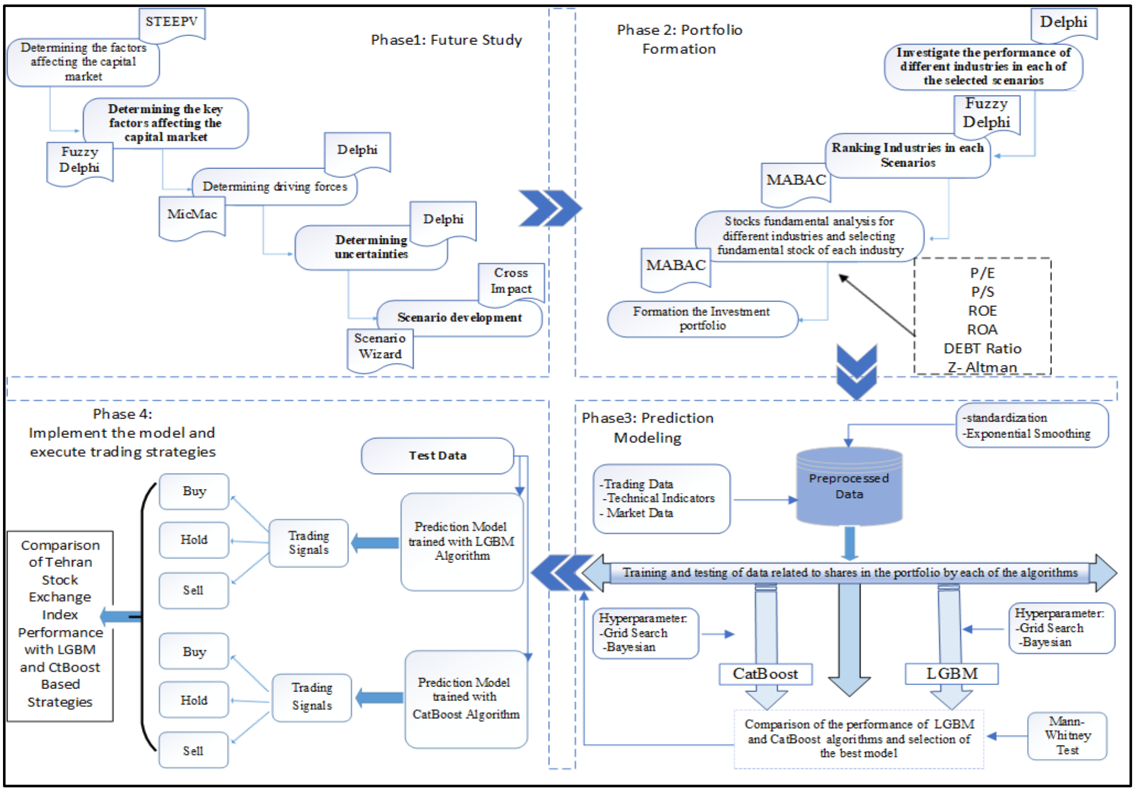

Figure 2: Conceptual model of research

Phase 1: Future Study of IRAN capital market

In order to develop different scenarios for the Iran capital market in the 2 years horizon, the following steps performed:

1) Determining the subject future study
   How can an investor gain more returns by identifying top industries in 2-Year horizon?
2) Determining the critical uncertainty
   In this stage with the participation of panel of expert and the use of environmental analysis and STEEPV matrix, the factors affecting the Tehran securities market are extracted. Then using Delphi method in the panel of expert and Delphi fuzzy method,
   the key factors affecting Tehran securities market are extracted.
3) Determining driving forces
   In this stage, we use cross impact analysis for determining future driving forces. For this purpose, the intensity of the effect of each key

factor on each other is determined using expert questionnaires and scoring in the range 0 to 6.

4) Determining the uncertainties and Trend Analysis:
At this stage, key uncertainties are determined for each of the driving factors using an expert questionnaire and impact chains of variables affecting each other calculated by Cross impact analysis.
5) Scenario Development:
At this stage, using GBN method and scenario wizard software, scenarios are developed and scenarios with the most impact factor and maximum consistency are selected.

Phase 2: Portfolio Formation

1) Review of different industries performance in each of the scenarios.
At this stage, using expert questionnaires, different industries are evaluated in terms of vulnerability to each of the scenarios and scored based on Likert scale.
2) Ranking industries in each scenario
The result of expert questionnaires was evaluated using fuzzy Delphi and ranked based on the least vulnerability with aid of multi criteria decision tool in facing all scenarios. After ranking, six industries with highest rank was selected.
3) Stocks fundamental analysis for different industries and selecting fundamental stock of each industry.
Selecting stocks in selected industries with more than 90% liquidity and more than 20% float and evaluating based on fundamental factors such as Altman Z, P/E, P/S, ROA, ROE and debt ratio.
4) Formation the Investment portfolio:
In this stage, each stock of 6 selected industries are ranked based on their fundamental status with use of multi criteria decision tools and one stock from each industry with highest fundamental rank is selected.

Phase 3: Modeling and Execute the Predictions

For Modeling the CRISP-DM data analysis (Shearer, 2000) methods is used according to following steps:

1) Industry analysis :(Business Understanding)
This section was done with use of future study tools and stock portfolio formation is the result of this stage.
2) Selecting model input and output variables (Data Preparation):
A. Input variable: including technical indicator, stock trading data (Opening, Closing Price, Last, Closing, High, Low and Volume), Stock market Index, Buying Power, P/S and P/E according to Table 2.
B. Output variables are selling and buy signals, which forecast the future price movement trend in a 15 days horizontal and 15 days forecasting windows. According to research of Shynkevich et al., 2017, the best performance is done with a forecast horizontal equal to forecasting windows.
3) Data preprocessing:
In this research, in addition to normalization, the data are exponential smoothed to remove past data noise (Trappecy, et.al, 2007)

4. Modeling

In this stage for each of the selected stocks 2 machine learning Forecasting algorithm called LGBM and CatBoost is used for modeling. For each algorithm price data and technical indicator is used. Data is processed in two categories training (80%), test (20%), and is trained for forecasting. After initial data training, with use of Bayesian hyper parameter for each model, the best value for the main parameters of the model selected and model inputs retrained with use of optimized model 5. Evaluating model results in cooperation with other models and set objectives.

After optimizing models and data training of each shares with use of optimized models. Model performance is compared using criteria of mean squares errors and accuracy of forecasting tools. The following Formula is used to check the accuracy of the forecasting tools: (5)

Accuracy= (True (+1) +True (-1)) / (True (+1) +True (-1) +False (+1) +False (-1)) Then the accuracy and performance of models is compared with each other's. Due to small numbers of stocks in portfolio (6 data for each model), Mann-Whitney test is used for comparison of LGBM and CatBoost Model accuracy. After accuracy comparison test, both LGBM and Catboost models is used for prediction and presenting trading signals for each stock in selected portfolios

Phase4: Model Deployment

1) Implementing the model and applying trading strategies

   In this research, it is assumed that at the March of 2019, all the stocks in selected portfolio are purchased at the weight assigned to it. The prediction is made for 15 future days and the stock kept in portfolio until the"-1" signal is received and repurchased as soon as "+1" signal is received. The mentioned process is continuing until the last received signal during the test period (18-Jan-2020). The predicted signals for the i^{the} day are calculated based on the following formula (Shynkevich, et al., 2017).

   $$Target_i = Sign\ (close_{i+d} - close_i) \quad (6)$$

   In above-mentioned formula, d is the number of days for which the prediction is made. $Close_i$ is closing price in i^{th} days. The value of "1" for the target means that there is an uptrend in the d next day. And "-1" means there is a downtrend in the next d days.

The trading strategy is also determined as follow:

IF TARGET =1 AND Portfolio Dose not have the Share THEN ➡ BUY
ELSEIF Portfolio Include the Share THEN ➡ HOLD
IF TARGET =-1 AND Portfolio Dose not have the Share THEN ➡ SALE
ELSEIF Portfolio Include the Share THEN ➡ HOLD

2) Calculate the performance of the selected portfolio based on each of the strategies:

   At this step, the selected portfolio return calculated after applying each of the buy& hold strategy and smart trading strategies during the period from March 2014 to the end of Jan 2020. Then is compared with each other and Market index.

   For calculating the return, the following formula is used: (7)

$$R_t = \sum (P_{t+d}*0.99 - P_t*1.005) / (p_t*1.005)$$

At above formula R_t is denote for each trading return, Pt+d denote selling price, Pt denote buying price. 0.5% is considered as buying commission and 1% is considered as selling commission.

Research main questions:

1) Do LGBM algorithm trading strategies perform better than CatBoost based trading strategies?
2) Do trading strategies based on active portfolio management (using machine-learning algorithms, fundamental and technical analysis) perform higher than the market index?

4. Research Results

4.1. Future Study and Fundamental analysis Results

At this stage, the key factors affecting the capital market were identified according to Table 3, for this purpose, using STEEPV matrix, multilayer analysis, review of studies and expert questionnaire; the factors affecting capital market are selected.

Then, by reviewing the factors affecting the capital market and determining the key factors in the Delphi panel, based on the results of fuzzy Delphi, the key factors are identified according to Table 3.

Table 3: Key Factors Affecting Iran Capital Market

Europe Leave Joint Comprehensive Plan of Action	Inflation Rate	Foreign Exchange Rate
The Result of INSTEX Review	Resumption of Iranian Oil Export	Iran leave Joint Comprehensive Plan of Action
Global Metal Price	Oil Price	Stability of Iranian Government
Extension of FATF Deadline	The Global Economic Prices	Global Gold Price
	Negotiation with USA	Iran's Banking System

Using the cross-impact method and the experts' questionnaire, the interaction of key factors on each other was determined. Using MicMac software, Figure 3 was extracted as Influence- Dependence Chart.The first quarter of this diagram introduces the drivers.

After Determining of uncertainties related to each driving forces and other effective factors, scenarios were developed using the scenario wizard software and 4 scenarios of Safe beach, Painkiller, Retaliate and Challenge with most impact factor and least inconsistency were selected according to Table (4). After reviewing, the results of the expert questionnaire using fuzzy Delphi method and MABAC tool, (Pamučar& Ćirović, 2015) industries were ranked in terms of vulnerability in the face of selected scenarios. Six pharmaceutical, metal, food, non-metallic Minerals, chemical and agricultural industries were selected with higher rank. 6 shares of Dekimi, Faravar, Ghegorji, Kafra, Shepaksa and Zemagsa which had the highest fundamental rating in their industries with MABAC rating, were selected as the top shares.

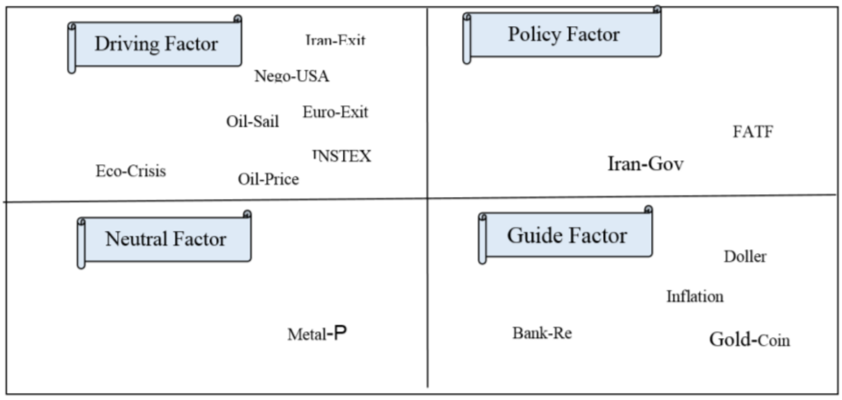

Figure3: Influence- Dependence Chart

After entering the key uncertainties in the scenario wizard software and determining the Cross impact of these uncertainties on each other, different scenarios resulting from these interactions are identified and four scenarios with the most impact factor and the least inconsistency are selected according to Table 4. Selected scenarios and the status of their driving forces and policymaking factors are shown in Table5.

After selecting the stocks with higher fundamental rank in each industry, 6 shares (Dekimi, Faravar, Ghegorji, Kafra, Shepaksa and Zemagsa according to Table 6. selected as the Best fundamental shares to form the medium terms portfolio

4.2. Prediction modeling results

After portfolio formation, both LGBM and CatBoost models train the data of each share. Result of Mann-Whitney test for model performance evaluation in Table 7. shows that in both LGBM and CatBoost models Bayesian optimization outperforms the grid search.

Comparison of LGBM base Strategies and CatBoost base strategies with Mann-Whitney Test in Table8. shows that the performance of trades, which is done base on LGBM signals, outperforms the CatBoost.

Finally, the performance of strategies, which are based on the forecast of LGMB and CatBoost algorithms were compared with each other. The result given in Table 9. Shows that The LGBM algorithm, with a return of 164%, has a better performance than CatBoost with a return of 142% and the Market index with a return of 138%.

Table4.Impact factor and Inconsistency of selected scenarios

Scenario / Specification	Challenge	Retaliate	Painkiller	Safe Beach
Impact Factor	38	7	27	48
Inconsistency	0	3	0	0

Table 5. Selected scenarios for the stock market in the 2-year horizon

Scenarios				Driving forces and policy making factors
Challenge	Retaliate	Painkiller	Safe Beach	
Exit	Reduction of obligation	No- Exit	No- Exit	IRAN Leave JCPOA
Exit	No- Exit	No- Exit	No- Exit	Europe Leave JCPOA
Not done	Not Done	Done	Done	INSTEX
Rejected	Suspension	Suspension	Passed	FATF
No negotiation	No negotiation	No negotiation	Negotiation	Negotiation with USA
Contractive Fiscal & Monetary	Contractive Monetary	Expansionary Monetary	Expansionary fiscal	Government monetary and Fiscal Policy
Increase	Minor Increase	Minor Decrease	Decrease	Inflation Rate
20-25	15 -20	10 - 12	8 - 9	Dollar Price in Iran
Trade War	Increase	Threat	Decrease	China-US Tensions
Disagreement	Increase	Negotiation Continuation	Decrease	UK-EU Tensions
Recession	Pessimism about the future	Normal	Relative Boom	Global Economic Crisis
Decrease	No Change	Decrease	Increase	Federal Reserve Interest Rate
1550-1650	1450-1500	1500-1550	1300 - 1400	Global Gold Price
Decrease	Minor Decrease	Normal	Increase	Global Metal Price
30-40 $	40-50 $	55-65 $	70-80 $	Global Oil Price

Table6. Fundamental Status of Selected Stocks

Raw	Industry	Share	W. F	ROE	ROA	P/S	P/E	Debt Ratio	Altman Z
1	pharmaceutical	Dekimi	22%	0.436	0.25	0.77	0.16	0.43	4.10
2	Metal	Faravar	20%	0.58	0.4	2.96	7.71	0.31	10.2
3	Food	Ghegorji	18%	0.4	0.22	1.55	9.36	0.54	4.82
4	non-metallic Minerals	Kafra	17%	0.58	0.33	2.19	5.98	0.43	6.25
5	Chemical	Shepaksa	14%	0.45	0.27	0.93	4.25	0.36	5.14
6	agricultural	Zemagsa	9%	0.41	0.11	1.51	8.83	0.66	1.84

Table7. Comparison of Bayesian and Grid Search hyperparameter

Model / Share	LGBM		CatBoost	
	Bayesian	Grid Search	Bayesian	Grid Search
Dekimi	93.49	78.69	84.02	76.33
Faravar	92.04	78.40	85.79	77.27
Ghegorji	96.8	84.8	96	95.2
Kafra	88	74.85	88.57	80
Shepaksa	89.73	77.83	87.02	77.83
Zemagsa	91.11	82.77	86.11	80
P-Value	0.004		0.037	

Table8. The Comparison on LGBM and CatBoost Based strategies

Model / Share	LGBM	CatBoost
Dekimi	93.49	84.02
Faravar	92.04	85.79
Ghegorji	96.8	96
Kafra	88	88.57
Shepaksa	89.73	87.02
Zemagsa	91.11	86.11
P-Value	0.049	

Table 9: Compaction of Different Strategies Return

Strategy / Share	Weight	LGBM	CatBoost	Tehran Overall Index
Dekimi	22	171	152	138%
Faravar	20	125	112	
Ghegorji	18	288	220	
Kafra	17	109	102	
Shepaksa	14	101	95	
Zemagsa	9	192	185	
Portfolio		164%	142%	

5. Discussion and Conclusions

The purpose of this study is to compare intelligent trading strategies based on LGBM and CatBoost algorithms in the stock market, Considering the opinions and decision-making methods of financial experts in all stages of designing the trading system.

For this purpose, using the future study in Delphi panel, and implementing decision making tools such as Delphi fuzzy and MABAC, four scenarios named Safe Beach, Pain Killer, Retaliate and Challenge with lowest inconsistency and highest impact factor was developed for 2 years horizontal of Iran capital market. Regarding impact of scenarios on all industries, industries with lowest vulnerability and highest return facing with whole scenarios selected. For each selected industries a fundamental analysis using P/E, P/S, Z-Altman, Debt to Asset ratio, ROE, and ROA criteria was performed and a fundamental stock portfolio was formed with less vulnerability to the occurrence of advanced scenarios in the Tehran Stock Exchange for the next two years. Then trading strategies which are based on optimized machine learning algorithms was used for trading on selected portfolio. In this regard, two prediction models developed by use of LGBM and CatBoost algorithms and optimized by use of Bayesian Hyperparameter.

The results show that the scenarios have the ability to define the stock portfolio and can identify industries with lower vulnerability and higher returns by making a medium-term forecast in the form of future study and considering political and economic factors affecting the capital market, which is in line with findings of (Hanafizadeh, et al., 2011)

On the other hand, the stock portfolio created by future study with has a higher return than the Tehran Stock Exchange Index, which is in line with findings of Roodpooshti and Shirin Bayan (2016).

In order to apply trading strategies, selecting stocks with a better fundamental specification in each industry can pave the way for higher stock portfolio performance, which is in line with findings of Shahmansouri (1396) research.

Comparison of prediction models using Mann-Whitney Test, which are based on the LGBM and CatBoost algorithms, shows that the performance of the LGBM model is higher than the CatBoost algorithm in terms of accurate detection of stock price movement trends and error reduction.

The results of using prediction models based on both LGBM and CatBoost algorithms for trading on the stock portfolio show that using trading strategies, which are based on signals generated by LGBM algorithm with 164% return, outperform CatBoost trading strategy with 142% and Tehran Stock Exchange Index with 138%.

These results, in terms of the success of active strategies based on the use of machine learning is in line with findings of most researchers such as Sarang et al. (2020), Rai and Hosseini (2015), Tehrani, Hindijani and Nowruzian (2015) and Bajelan, et.al. (2016).

The results of optimizing the models by Bayesian hyper parameters using Mann-Whitney Test with P-Value of 0.004 for LGBM and P-Value of 0.037 for CatBoost indicate an increase in the accuracy of prediction relative to the Greed Search hyper parameter, which is in line with findings of Yufi, et.al (2017) research.

Table 10: Summary of research results

Research Question	Result
Do scenarios have the ability to define the stock portfolio with lower vulnerability and higher returns?	A fundamental stock portfolio was formed with less vulnerability to the occurrence of advanced scenarios in the Tehran Stock Exchange for the next two years
Do LGBM algorithm trading strategies perform better than CatBoost based trading strategies?	Trading strategies based on LGBM algorithm signal with 164% return outperformed the CatBoost with 142%.
Do trading strategies based on active portfolio management (using Future Study, machine-learning algorithms, fundamental analysis) perform higher than the Tehran Stock market index?	Trading strategies based on active portfolio management with 164% outperformed Tehran Stock market index with 138%.
Do the model optimized with Bayesian hyperparameter outperforms the models optimized with Greed search?	The Result of Mann-Whitney test indicate better performance of Bayesian hypermeter

It is suggested for future research:

1) An ensemble models, consist of boosting models such as XG-Boost and deep learning algorithms can be used for prediction modeling.
2) Experts technical trading rules such as Elliott waves and Ichimoku can be used as input variables of prediction models.
3) Other methods of future study such as Assumption-Based Planning can be used.
4) Stock data such as Smart Money can be use as input variable.

References

* Atsalakis, G.S., Valavanis, K.P., (2009). Surveying stock market forecasting techniques part II: soft computing methods, Expert Syst. Appl.36 (3), 5932-5941.
* Bajelan, Saeed, Fallahpour, Saeed, Dana,Shirin.(2016), The Stock Trend Prediction Using Volume Weighted Support Vector Machine with a Hybrid Feature Selection Method to Predict the Stock Price Trend in Tehran Stock Exchange, Financial Management Strategy,4(3),121-148.
* Choudry,R, Grag, K. (2008). A Hybrid Machin Learning System for Stock Market Forecasting.Word Academy of Science, Engineering and Technology, 39.
* Esfandyar Shahmansuri,(2017), Testing portfolio strategies based on fundamental, technical and heuristically with behavioral objects and characteristics of Tehran stoke market investors, Investment Knowledge,6(24),99-116.
* Fama, E.-F., Blume, M.-E., (1966), Filter rules and stock market trading, J. Bus. 39 (1), 226–241.
* gholamian, elham, davoodi , sayyed mohammad reza,(2018), Predicting the Direction of Stock Market Prices Using Random Forest, financial engineering and portfolio management,9(35), 301-322.
* Gulin, Ke, et.al, (2017), LightGBM: A Highly Efficient Gradient Boosting Decision Tree, Neural Information Processing Systems, 30.
* Hanafizadeh, P., Kazazi, A., and Jalili Bolhasani, A., (2011) Portfolio design for investment companies through scenario planning, Management Decision, 49(4), 513-532.
* Huang, Gumine, et.al, (2019), Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions, Hydrology,574,1029-1041.
* Jing, Zhang, et.al. (2018), A novel data- driven stock price trend prediction system, Expert system with applications, 97,60 - 69.
* Lawrence, R. (1997), Using Neural Networks to Forecast Stock Market Prices, 1-12.
* Leung, M, T., Chen, A. S., & Daouk, H. (2000). Forecasting exchange rates general regression neural networks.
* Moshari, Mohammad,et.al,(2019), Designing a Hybrid Intelligent Model for Prediction of Stock Price Golden Points, Investment Knowledge,8(29),45-66.
* Nanni, L., & Lumini, A. (2009). An experimental comparison of ensemble of classi- fiers for bankruptcy prediction and credit scoring. Expert Systems with Applica- tions, 36, 3028–3033.
* Pamučar, Ddragan, Ćirović, Goran, (2015), the selection of transport and handling resources in logistics centres using Multi-Attributive Border

Approximation Area Comparison (MABAC), Expert system with applications, 42, 3016- 3028.

* Parsa M, et.al. (2020) Bayesian Multi-objective hyperparameter Optimization for accurate, Fast, and Efficient Neural, Frontiers in Neuroscience,14. Network Accelerator Design.
* Patel, J, et.al. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machinelearnning techniques. Expert Systems with Applications, 42 (1), 259–268.
* Raei, Reza, Hoseini, Farhang,(2015),Comparison of trading returns based on technical indicators and fuzzy logic and the combined method of genetic algorithm - fuzzy logic, financial engineering and portfolio management,6(24)
* Rahnamay Roodposhti, feridoon, Shirinbayan ,Neda,(2016), Design portfolio using a scenario planning approach using Assumption-based planning, financial engineering and portfolio management,7(28), 1-21.
* Saranj, et.al. (2020). Developing a Stock Technical Trading System Integrating MLP Neural Network with Evolutionary Algorithms, Financial knowledge and securities Analysis, 13(45), 47-64.
* Shearer C. (2000), The CRISP-DM model: the new blueprint for data mining, J Data Warehousing (2000); 5:13-22.
* Sarkar, Sobhan, et al. (2019), Application of optimized machine learning techniques for prediction of Occupational accidents,Computer & Operation research, 106, 210-224.
* Shynkevich, Yauheniya, et al. (2017), forecasting price movements using technical indicators: Investigating the impact of varying input window length. Neurocomputing, 264 71–88.
* Suryoday, Basak, et al., (2018), predicting the direction of stock market prices using tree-based classifiers, North American Journal of Economics and finance, In Press.
* Tehrani, Reza, Hendijanizadeh, Mohammad, Noruzian Lakvan, Eisa,(2015), A novel approach to fulfill active portfolio management and automatic stock trading based on feature selection algorithm, Investment Knowledge,4(13),107-126
* Trappey, C., Shin, T. and Trappey, A. (2007), "Modeling international investment decisions for financial holding companies", European Journal of Operational Research, Vol. 180, pp. 800-14.
* Xiaolei, Sun, Mingxi, Liu, Zeqian, Sima, (2020). A novel cryptocurrency price trend forecasting model base on LightGBM, Finance Research Letters, 32.
* Xiao-dan Z., Ang L., Ran P., (2016), Stock trend prediction based on a new status box method and AdaBoost probabilistic support vector machine, Journal of Applied Soft Computing, 385-398.
* Yufei, Xia, Chuanzhe, liu, YuYing, Li, Nana, Liu, (2017). Aboosted decision tree approach using Bayesian hyper_parameter optimization for credit scoring, Expert System with Application, 78, 225-241.
* Wu, Jia, et.al, (2019), Hyper parameter Optimization for Machine Learning Models Based on Bayesian Optimization, Electronic Science and Technology, 17(1).